computers

MDPI

Article

On Granular Rough Computing: Handling Missing Values by Means of Homogeneous Granulation †

Piotr Artiemjew *,‡ **and Krzysztof Ropiak** ‡

Faculty of Mathematics and Computer Science, University of Warmia and Mazury in Olsztyn, 10-710 Olsztyn, Poland; kropiak@matman.uwm.edu.pl

* Correspondence: artem@matman.uwm.edu.pl

† Extended version of paper "Missing values absorbtion based on homogeneous granulation" presented at the 25th International Conference on Information and Software Technologies (ICIST 2019) held on 10–12 October 2019 in Vilnius, Lithuania.

‡ These authors contributed equally to this work.

Received: 2 January 2020; Accepted: 12 February 2020; Published: 15 February 2020

Abstract: This paper is a continuation of works based on a previously developed new granulation method—homogeneous granulation. The most important new feature of this method compared to our previous ones is that there is no need to estimate optimal parameters. Approximation parameters are selected dynamically depending on the degree of homogeneity of decision classes. This makes the method fast and simple, which is an undoubted advantage despite the fact that it gives a slightly lower level of approximation to our other techniques. In this particular article, we are presenting its performance in the process of missing values absorption. We test selected strategies on synthetically damaged data from the UCI repository. The added value is to investigate the specific performance of our new granulation technique in absorbing missing values. The effectiveness of their absorption in the granulation process has been confirmed in our experiments.

Keywords: granular rough computing; missing values handling; homogeneous granulation

1. Introduction

Granular computing is a paradigm, dedicated to computing, based on objects similar to each other on the basis of selected similarity measure. The idea was proposed by Lotfi Zadeh [1,2]. Granulation is a part of the fuzzy theory by the very definition of fuzzy set, where inverse values of fuzzy membership functions are the basic forms of granules. Shortly after Lotfi Zadeh proposed the idea of granular computing, the granules were introduced in terms of rough set theory by T.Y. Lin, L. Polkowski, and A. Skowron. In rough set theory, granules are defined as classes of indiscernibility relations. Interesting research on more flexible granules based on blocks was conducted by (Grzymala–Busse), and templates by (H.S. Nguyen). The granules based on rough inclusions was introduced by (Polkowski and Skowron [3]), based on tolerance or similarity relations, and, more generally, binary relations by (T.Y. Lin [4], Y. Y. Yao [5–7]). Being in the context of rough mereology was proposed by L. Polkowski and A.Skowron, approximation spaces by A. Skowron and J. Stepaniuk [8,9], and logic for approximate reasoning by L.Polkowski and M. Semeniuk-Polkowska [10], and Qing Liu [11]. Examples of interesting studies from recent years can be found in [12–18].

This is a work about using granular rough computing techniques to absorb missing values [19]. The exact theoretical introduction to the family of approximation methods to which our methods belong to can be found in [20–22]. Of course, to understand the body of the algorithmic work, we have included all the relevant details in the following sections.

Our recently developed homogeneous granulation method is described in [23]. The main difference in our granulation algorithm, to the previously developed ones, is that there is no need to estimate the granulation radius. This parameter is being set automatically depending on the indiscernibility level of the decision classes. The degree of indiscernibility is the percentage of attributes for which objects are identical. Homogeneous granulation was also implemented in the epsilon variant for numerical data, which was described in [24–26] as well as being a part of a novel ensemble model—Ensemble of granular reflections—described in detail in [27]. The main motivation to carry out the tests were our previous research results in the context of absorption of unknown values, which gave very interesting results. The creation of a new technique naturally caused scientific curiosity to examine its performance in the same context.

This paper presents some preliminary experiments of using the homogeneous granulation as a missing values absorption technique. We have taken into consideration four absorption strategies, which we have named *A*, *B*, *C*, and *D*. Results of using those strategies with other granulation methods are available in Polkowski and Artiemjew [28,29].

Below is a detailed description of the strategies used.

1.1. Selected Key Variants

The strategies to consider are as follows:

1. *Variant A: in granulation process* $*$*=each value, when fixing the unadsorbed values* $*$, $*$*=each value.*
2. *Variant B: in granulation process* $*$*=each value, when fixing the unadsorbed values* $*, * = *$.
3. *Variant C: in granulation process* $* = *$*, when fixing the unadsorbed values* $*$, $*$*=each value.*
4. *Variant D: in granulation process* $* = *$*, when fixing the unadsorbed values* $*, * = *$.

In the granulation process, taking *A* and *B* strategies into consideration, stars evaluation of the similarity to any other value is always positive. For *C* and *D* variants, stars are treated as stars—so they evaluate as positive only when comparing with other stars. Those strategies bring up the following granulation definition:

In the following variants, we see the process of granule formation, where we use two basic options when comparing descriptors $* = each\ value$ and $* = *$. Obviously, in the $* = each\ value$ variant, the granules increase their size, i.e., after approximation they absorb potentially more damage values.

Considering ob_1 as the center of the granule, ob_2 as the compared object of the training system, *r* as the indiscernibility degree of the descriptors, IND as the indiscernibility relation, and *d* as the decision attribute, we have considered the following options of internal processes for repairing unknown values. For readability, we have placed the legend of the applied markings in Table 1.

Table 1. The legend of the applied markings.

Name	Description
$TRAIN_i$	*i*-th training decision system, used in cross validation process
ob_i	*i*-th object of selected decision system
gran	granule
radius	granulation radius
Attr	set of conditional attributes
IND	indiscernibility relation
d	decision attribute
MaVot	Majority Voting procedure
conc_dep	concept-dependent variant of granulation
$\|set\|$	cardinality of the set

1.1.1. For Variant $* = each\ Value$, the Granulation Process of the ***i*-th Training** Set $TRAIN_i$ Looks as Follows (A, B Variants)

$$gran_{radius}^{conc\_dep,*=each\ value}(ob_1) = \{ob_2 \in TRAIN_i : \frac{|IND^{*=each\ value}(ob_1, ob_2)|}{|Attr|} \leq radius\ \&\ d(ob_1) = d(ob_2)\},$$

for IND defined as

$$IND^{*=each\ value}(ob_1, ob_2) = \{a \in Attr : a(ob_1) = a(ob_2) \parallel a(ob_1) = * \parallel a(ob_2) = *\}.$$

where & means AND, $\parallel$ means OR.

1.1.2. For Variant $* = *$, The Granulation Process of the Set $TRAIN_i$ Looks as Follows (C, D Variants)

$$gran_{radius}^{conc\_dep,*=*}(ob_1) = \{ob_2 \in TRAIN_i : \frac{|IND^{*=*}(ob_1, ob_2)|}{|Attr|} \leq radius\ \&\ d(ob_1) = d(ob_2)\},$$

for IND defined as

$$IND^{*=*}(ob_1, ob_2) = \{a \in Attr : a(ob_1) = a(ob_2)\}.$$

1.1.3. For Variant $* = each\ Value$, the Way We are Fixing the Unadsorbed Values of $TRAIN_i$ Looks as Follows (A, C Variants)

In this variant, the saved stars are replaced with the most frequently appearing attribute value of the training system.

For variant A, granule surrounding the defective sample $MaVot(gran_{radius}^{conc\_dep,*=each\ value}(ob_1))$ (further mark as $temp$) looks like the below:

$$\text{if } a_j(temp) = *,$$

The repairing process looks as follows:

$$gran_{radius,a_j}^{conc\_dep,*=each\ value}(temp) = \{ob_2 \in TRAIN_i : \frac{|IND_{a_j}^{*=each\ value}(temp, ob_2)|}{|Attr|} \leq radius\ \&\ d(temp) = d(ob_2)\},$$

IND is defined as

$$IND_{a_j}^{*=each\ value}(temp, ob_2) = \{a \in Attr : (a(temp) = a(ob_2) \parallel a(temp) = * \parallel a(ob_2) = *)\ \&\ a_j(ob_2)! = *\}.$$

For variant C, granule surrounding the defective sample $MaVot(gran_{radius}^{conc\_dep,*=*}(ob_1))$ (further mark as $temp2$) looks like the below:

$$\text{if } a_j(temp2) = *,$$

The repairing process looks as follows:

$$gran_{radius,a_j}^{conc\_dep,*=each\ value}(temp2) = \{ob_2 \in TRAIN_i : \frac{|IND_{a_j}^{*=each\ value}(temp2, ob_2)|}{|Attr|} \leq radius\ \&\ d(temp2) = d(ob_2)\},$$

IND is defined as

$$IND_{a_j}^{*=each\ value}(temp2, ob_2) = \{a \in Attr : (a(temp2) = a(ob_2) \parallel a(temp2) = * \parallel a(ob_2) = *) \,\&\, a_j(ob_2)! = *\}.$$

1.1.4. For Variant $* = *$, the Way We Fix the Unabsorbed Values of $TRAIN_i$ Looks as Follows (B, D Variants)

In this variant, the saved stars are completed with the most frequently appearing attribute value of the training system.

For variant B, granule surrounding the defective sample $MaVot(gran_{radius}^{conc\_dep,*=each\ value}(ob_1))$ (further mark as $temp3$) can be defined as follows:

$$gran_{radius,a_j}^{conc\_dep,*=*}(temp3) = \{ob_2 \in TRAIN_i : \frac{|IND_{a_j}^{*=*}(temp3, ob_2)|}{|Attr|} \leq radius \,\&\, d(temp3) = d(ob_2)\},$$

IND is defined as

$$IND_{a_j}^{*=*}(temp3, ob_2) = \{a \in Attr : a(temp3) = a(ob_2) \,\&\, a_j(ob_2)! = *\}.$$

For variant D, the granule surrounding the defective sample $MaVot(gran_{radius}^{conc\_dep,*=*}(ob_1))$ (further mark as $temp4$) looks like that shown below:

$$gran_{radius,a_j}^{conc\_dep,*=*}(temp4) = \{ob_2 \in TRAIN_i : \frac{|IND_{a_j}^{*=*}(temp4, ob_2)|}{|Attr|} \leq radius \,\&\, d(temp4) = d(ob_2)\},$$

IND is defined as

$$IND_{a_j}^{*=*}(temp4, ob_2) = \{a \in A : a(temp4) = a(ob_2) \,\&\, a_j(ob_2)! = *\}.$$

2. Homogenous Granulation in $* = *$ and $* = each\ value$ Variant

For $IND^{*=each\ value}(ob_1, ob_2)$, and variant $* = each\ value$, we create the granule,

$$gran_{r_u}^{homogenous,*=each\ value} = \{ob_2 \in U : |gran_{r_u}^{conc\_dep,*=each\ value}| - |gran_{r_u}^{*=each\ value}| == 0,$$

$$for\ minimal\ r_u\ fulfills\ the\ equation\}$$

$$gran_{r_u}^{conc\_dep,*=each\ value} = \{ob_2 \in U : \frac{IND^{*=each\ value}(ob_1, ob_2)}{|Attr|} \leq r_u \,\&\, d(ob_1) == d(ob_2)\}$$

$$gran_{r_u}^{*=each\ value} = \{ob_2 \in U : \frac{IND^{*=each\ value}(ob_1, ob_2)}{|Attr|} \leq r_u\}$$

$$r_u = \{\frac{i}{|Attr|},\ where\ i = 0., 1., ..., |Attr|\}$$

in case of $* = *$ variant and earlier defined $IND^{*=*}(ob_1, ob_2)$, the granule is as follows:

$$gran_{r_u}^{homogenous,*=*} = \{ob_2 \in U : |gran_{r_u}^{conc\_dep,*=*}| - |gran_{r_u}^{*=*}| == 0,$$

$$for\ minimal\ r_u\ fulfills\ the\ equation\}$$

$$gran_{r_u}^{conc\_,*=*} = \{ob_2 \in U : \frac{IND^{*=*}(ob_1, ob_2)}{|Attr|} \leq r_u \,\&\, d(ob_1) == d(ob_2)\}$$

$$gran_{r_u}^{*=*} = \{ob_2 \in U : \frac{IND^{*=*}(ob_1, ob_2)}{|Attr|} \leq r_u\}$$

$$r_u = \{\frac{i}{|Attr|}, \ where \ i = 0., 1., ..., |Attr|\}$$

3. Testing Session

This section is describing the experimental part followed by presentation of the results. The effectiveness was calculated on an artificially damaged datasets (10 percent of the data has been replaced with stars) chosen from UCI Repository [30].

3.1. The Steps of the Procedure

(i) Selected dataset was uploaded,
(ii) The data have been prepared for the Cross Validation 5 model,
(iii) The $TRAIN_i^{complete}$ was granulated using a proper variant,
(iv) The $TEST_i$ was classified based on $TRAIN_i^{complete}$ using kNN (the nil case),
(v) $TRAIN_i^{complete}$ was filled with a fixed percentage of random stars;
(vi) $TRAIN_i$ was fixed by granulation based on the chosen variant—A, B, C, or D
(vii) classification of $TEST_i$ was performed based on a fixed training system using kNN,
(viii) the average result of the classification was calculated from all of the folds,

We perform the procedure five times, receiving the mean value from each test 5$Cross_V5$).

3.2. Verification of Results Stability

We have computed an additional parameter to show the bias of accuracy, defined as follows:

$$Bias\_Acc = \frac{\sum_{i=1}^{5}(max(accuracy_1^{Cross\_V5}, accuracy_2^{Cross\_V5}, ..., accuracy_5^{Cross\_V5}) - accuracy_i^{Cross\_V5})}{5}, \quad (1)$$

for

$$accuracy = \frac{\sum_{i=1}^{5} accuracy_i^{Cross\_V5}}{5}.$$

The classifier used for our experiments is a classical kNN, where the smallest summary distance of k-nearest objects indicates the decision parameter value. k parameters are estimated with the $Cross\_V5$ method on a sample of data, which resulted in $k = 5$ for Australian Credit. $k = 3$ for Pima Indians Diabetes, $k = 19$ for Heart Disease, $k = 3$ for Hepatitis, and $k = 18$ for the German Credit data set. We have selected the kNN classifier for testing due to the fact that, in past tests, testing other granulation variants to absorb unknown values, we used the same classification variant as the base classifier. Our performance tests, NB classifier, kNN, SVM, and deep neural networks, showed that kNN is fully comparable with the best classifiers in the context of granular reflection based classification.

3.3. Overview of the Testing Results

The results of missing values absorption using concept dependent granulation are shown in Tables 2–6. For homogeneous granulation, please refer to Tables 7–11. As a conclusion of the research presented in [22], we can say that granulation is an effective technique of absorbing some degree of missing values placed in the dataset. Our observations were proved by comparable classification results with the non-missing values data case. We need to point out that granulation brings another important benefit—it can significantly (up to 80 percent) reduce the number of objects used for classification. As shown in [22], this behavior strictly depends on the diversity of used datasets. Using strategies A and B for lower values of granulation radius, the approximation is faster because the $* = each\ value$ variant causes a higher number of objects in the granules. In case of $* = *$, stars can increase diversity

of the data and consequently a higher number of granules containing fewer number of objects than in the $*$ = *each value* case.

Table 2. Missing values handling using *conc_dep* granulation technique; $5 \times Cross\_V5$; A, B, C, D variants vs. nil case (classification based on original, undamaged training system); *Australian Credit*; *synthetic* 10% *damage*; *radius* = indiscernibility ratio; *Bias_Acc* = defined in Equation (1); *Gran_Size* = the number of training objects after granulation.

(**a**)

	Accuracy					**Bias_Acc**				
radius	***nil***	***A***	***B***	***C***	***D***	***nil***	***A***	***B***	***C***	***D***
0	0.772	0.773	0.773	0.773	0.773	0.002	0.005	0.005	0.005	0.005
0.0714286	0.772	0.773	0.773	0.772	0.773	0.002	0.005	0.005	0.006	0.006
0.142857	0.77	0.773	0.772	0.773	0.773	0.012	0.005	0.006	0.01	0.01
0.214286	0.79	0.776	0.777	0.805	0.795	0.01	0.011	0.013	0.026	0.009
0.285714	0.798	0.777	0.778	0.812	0.808	0.012	0.015	0.017	0.017	0.008
0.357143	0.815	0.783	0.778	0.829	0.829	0.018	0.017	0.015	0.008	0.008
0.428571	0.837	0.788	0.794	0.842	0.837	0.016	0.008	0.003	0.012	0.006
0.5	0.838	0.82	0.818	0.841	0.848	0.011	0.008	0.017	0.01	0.016
0.571429	0.847	0.831	0.824	0.846	0.849	0.011	0.012	0.022	0.01	0.006
0.642857	0.849	0.839	0.835	0.852	0.848	0.014	0.014	0.009	0.005	0.007
0.714286	0.851	0.836	0.833	0.855	0.85	0.008	0.011	0.013	0.006	0.006
0.785714	0.858	0.837	0.84	0.846	0.852	0.013	0.008	0.01	0.012	0.013
0.857143	0.861	0.849	0.848	0.848	0.849	0.013	0.008	0.007	0.02	0.011
0.928571	0.863	0.849	0.847	0.848	0.85	0.011	0.012	0.012	0.011	0.011
1	0.862	0.849	0.849	0.85	0.85	0.012	0.008	0.008	0.011	0.011

(**b**)

	Gran_Size				
radius	***nil***	***A***	***B***	***C***	***D***
0	2	2	2	2	2
0.0714286	2.48	2	2	2.92	2.84
0.142857	3.6	2.12	2.16	4.56	4.48
0.214286	5.08	2.88	2.88	8.6	8.12
0.285714	8.44	4.24	4.28	15.36	15.52
0.357143	15.28	6.4	6.2	32.88	33.16
0.428571	32.24	9.16	9.88	70.12	70.08
0.5	70.04	18.4	17.88	148.8	148.48
0.571429	157.76	33.4	34.68	283.36	283.28
0.642857	318.04	73.44	73.64	431.72	431.56
0.714286	467.12	165	163.6	520.72	521.04
0.785714	536.08	322.56	321.24	546.76	546.72
0.857143	547.16	469.64	469.96	550.76	550.8
0.928571	548.72	536.48	536.52	551.8	551.8
1	552	550.56	550.56	552	552

Table 3. Missing values handling using *conc_dep* granulation technique; 5 × *Cross_V*5; A, B, C, D variants vs. nil case (classification based on original, undamaged training system); *Pima Indians Diabetes*; *synthetic* 10% *damage*; *radius* = indiscernibility ratio; *Bias_Acc* = defined in Equation (1); *Gran_Size* = the number of training objects after granulation.

(**a**)

	Accuracy					Bias_Acc				
radius	***nil***	***A***	***B***	***C***	***D***	***nil***	***A***	***B***	***C***	***D***
0	0.598	0.606	0.606	0.606	0.606	0.008	0.014	0.014	0.014	0.014
0.125	0.598	0.601	0.607	0.586	0.601	0.024	0.005	0.015	0.006	0.015
0.25	0.621	0.598	0.611	0.622	0.626	0.018	0.027	0.028	0.005	0.01
0.375	0.644	0.606	0.594	0.647	0.645	0.026	0.022	0.03	0.023	0.006
0.5	0.647	0.591	0.581	0.64	0.64	0.01	0.029	0.055	0.028	0.006
0.625	0.649	0.6	0.595	0.64	0.64	0.004	0.051	0.038	0.01	0.006
0.75	0.651	0.633	0.63	0.633	0.636	0.006	0.012	0.024	0.006	0.014
0.875	0.651	0.637	0.638	0.637	0.636	0.006	0.011	0.011	0.008	0.014
1	0.651	0.636	0.636	0.636	0.636	0.006	0.014	0.014	0.014	0.014

(**b**)

	Gran_Size				
radius	***nil***	***A***	***B***	***C***	***D***
0	2	2	2	2	2
0.125	34.2	3.2	3.16	32.56	31.96
0.25	154.44	9.24	8.44	146.92	147.44
0.375	365.8	30.4	28.96	363.48	363.28
0.5	539.36	94.12	90.16	546.92	547.08
0.625	609.92	250.08	248.36	610.08	610.12
0.75	614.4	485.6	490.72	614.24	614.24
0.875	614.4	597.48	598.6	614.4	614.4
1	614.4	613.48	613.48	614.4	614.4

Table 4. Missing values handling using *conc_dep* granulation technique; 5 × *Cross_V*5; A, B, C, D variants vs. nil case (classification based on original, undamaged training system); *Heart disease*; *synthetic* 10% *damage*; *radius* = indiscernibility ratio; *Bias_Acc* = defined in Equation (1); *Gran_Size* = the number of training objects after granulation.

(**a**)

	Accuracy					Bias_Acc				
radius	***nil***	***A***	***B***	***C***	***D***	***nil***	***A***	***B***	***C***	***D***
0	0.787	0.787	0.787	0.787	0.787	0.016	0.021	0.021	0.021	0.021
0.0769231	0.787	0.787	0.787	0.789	0.789	0.016	0.021	0.021	0.019	0.019
0.153846	0.788	0.787	0.787	0.794	0.794	0.019	0.021	0.021	0.013	0.013
0.230769	0.798	0.792	0.791	0.809	0.811	0.01	0.016	0.016	0.013	0.019
0.307692	0.807	0.79	0.787	0.813	0.815	0.012	0.021	0.017	0.02	0.015
0.384615	0.827	0.793	0.798	0.823	0.823	0.006	0.01	0.017	0.01	0.007
0.461538	0.824	0.813	0.81	0.821	0.817	0.013	0.016	0.019	0.008	0.005
0.538462	0.834	0.804	0.812	0.825	0.826	0.007	0.003	0.01	0.016	0.004
0.615385	0.823	0.82	0.821	0.827	0.831	0.014	0.024	0.016	0.006	0.006
0.692308	0.833	0.819	0.819	0.831	0.827	0.012	0.018	0.018	0.006	0.013
0.769231	0.829	0.823	0.823	0.832	0.83	0.004	0.01	0.01	0.009	0.007
0.846154	0.829	0.827	0.827	0.828	0.83	0.008	0.01	0.014	0.013	0.007
0.923077	0.829	0.829	0.829	0.83	0.83	0.008	0.008	0.008	0.007	0.007
1	0.829	0.83	0.83	0.83	0.83	0.008	0.007	0.007	0.007	0.007

(**b**)

	Gran_Size				
radius	***nil***	***A***	***B***	***C***	***D***
0	2	2	2	2	2
0.0769231	2.04	2	2	2.76	2.76
0.153846	3.16	2.16	2.2	4.56	4.56
0.230769	4.76	2.8	2.8	8	8
0.307692	8.96	3.8	3.76	15.12	15.12
0.384615	16.64	6	5.92	30.4	30.28
0.461538	34.44	11.16	11.24	60.68	60.56
0.538462	70.12	20.12	20.12	111.76	111.76
0.615385	127.32	38.4	38.4	168.68	168.68
0.692308	181.16	78.84	79.2	204.12	204.12
0.769231	210.56	142.16	142.16	214.56	214.56
0.846154	216	192.44	192.44	215.96	215.96
0.923077	216	212.72	212.72	216	216
1	216	215.64	215.64	216	216

Table 5. Missing values handling using *conc_dep* granulation technique; 5 × *Cross_V*5; *A*, *B*, *C*, *D* variants vs. nil case (classification based on original, undamaged training system); *Hepatitis*; *synthetic* 10% *damage*; *radius* = indiscernibility ratio; *Bias_Acc* = defined in Equation (1); *Gran_Size* = the number of training objects after granulation.

(**a**)

	Accuracy					**Bias_Acc**				
radius	***nil***	***A***	***B***	***C***	***D***	***nil***	***A***	***B***	***C***	***D***
0	0.817	0.822	0.822	0.822	0.822	0.022	0.017	0.017	0.017	0.017
0.0526316	0.817	0.822	0.822	0.822	0.822	0.022	0.017	0.017	0.017	0.017
0.105263	0.817	0.822	0.822	0.822	0.822	0.022	0.017	0.017	0.017	0.017
0.157895	0.817	0.822	0.822	0.822	0.822	0.022	0.017	0.017	0.017	0.017
0.210526	0.817	0.822	0.822	0.823	0.823	0.022	0.017	0.017	0.022	0.022
0.263158	0.817	0.822	0.822	0.83	0.83	0.022	0.017	0.017	0.015	0.015
0.315789	0.825	0.822	0.822	0.843	0.843	0.021	0.017	0.017	0.015	0.015
0.368421	0.823	0.825	0.825	0.843	0.843	0.009	0.021	0.021	0.015	0.015
0.421053	0.836	0.826	0.823	0.857	0.859	0.022	0.019	0.022	0.046	0.044
0.473684	0.868	0.841	0.84	0.872	0.871	0.009	0.017	0.012	0.025	0.026
0.526316	0.863	0.852	0.849	0.883	0.89	0.008	0.013	0.015	0.021	0.026
0.578947	0.877	0.859	0.855	0.874	0.879	0.019	0.031	0.028	0.01	0.012
0.631579	0.883	0.871	0.863	0.885	0.872	0.021	0.026	0.014	0.018	0.025
0.684211	0.889	0.883	0.877	0.885	0.879	0.008	0.008	0.013	0.025	0.025
0.736842	0.881	0.892	0.885	0.893	0.881	0.015	0.031	0.037	0.023	0.028
0.789474	0.893	0.885	0.89	0.898	0.88	0.004	0.05	0.045	0.025	0.017
0.842105	0.892	0.879	0.868	0.893	0.886	0.005	0.018	0.022	0.023	0.017
0.894737	0.892	0.876	0.883	0.875	0.883	0.005	0.021	0.014	0.035	0.034
0.947368	0.892	0.875	0.876	0.884	0.884	0.005	0.022	0.027	0.026	0.019
1	0.892	0.884	0.884	0.884	0.884	0.005	0.019	0.019	0.019	0.019

(**b**)

	Gran_Size				
radius	***nil***	***A***	***B***	***C***	***D***
0	2	2	2	2	2
0.0526316	2	2	2	2	2
0.105263	2	2	2	2.04	2.04
0.157895	2.08	2	2	2.24	2.24
0.210526	2.32	2	2	3.08	3.08
0.263158	2.72	2.12	2.12	4.32	4.32
0.315789	3.44	2.24	2.24	6.24	6.24
0.368421	5.24	2.96	3	9.6	9.6
0.421053	7.48	3.76	3.76	15.52	15.52
0.473684	11.72	5	5	24.88	24.88
0.526316	19.28	7.56	7.56	38.52	38.52
0.578947	30.48	11.96	11.96	58.24	58.24
0.631579	47.68	18.28	18.48	79.8	79.8
0.684211	69.96	28.72	28.72	99.4	99.4
0.736842	90	46.52	46.56	112	112
0.789474	109.48	69.28	69.28	119.2	119.2
0.842105	116.96	94.2	94.2	122.48	122.48
0.894737	121	111.32	111.36	123.56	123.56
0.947368	121.96	119.84	119.8	123.96	123.96
1	124	123.36	123.36	124	124

Table 6. Missing values handling using *conc_dep* granulation technique; 5 × *Cross_V*5; *A*, *B*, *C*, *D* variants vs. nil case (classification based on original, undamaged training system); *German credit*; *synthetic* 10% *damage*; *radius* = indiscernibility ratio; *Bias_Acc* = defined in Equation (1); *Gran_Size* = the number of training objects after granulation.

(**a**)

	Accuracy					**Bias_Acc**				
radius	***nil***	***A***	***B***	***C***	***D***	***nil***	***A***	***B***	***C***	***D***
0	0.564	0.57	0.57	0.57	0.57	0	0.012	0.012	0.012	0.012
0.05	0.564	0.57	0.57	0.57	0.57	0	0.012	0.012	0.012	0.012
0.1	0.564	0.57	0.57	0.57	0.57	0	0.012	0.012	0.012	0.012
0.15	0.564	0.57	0.57	0.58	0.58	0	0.012	0.012	0.01	0.01
0.2	0.569	0.57	0.569	0.585	0.58	0.001	0.012	0.01	0.005	0.007
0.25	0.584	0.57	0.569	0.606	0.604	0.002	0.012	0.01	0.01	0
0.3	0.617	0.578	0.583	0.647	0.649	0.006	0.011	0.007	0.004	0.003
0.35	0.647	0.585	0.583	0.673	0.674	0.028	0.016	0.013	0.006	0.002
0.4	0.657	0.597	0.598	0.692	0.692	0.008	0.003	0.002	0.006	0.01
0.45	0.696	0.64	0.635	0.687	0.682	0.002	0.014	0.014	0.003	0.002
0.5	0.7	0.664	0.657	0.716	0.71	0.003	0.008	0.009	0.009	0.002
0.55	0.698	0.64	0.639	0.718	0.716	0.006	0.018	0.016	0.005	0.016
0.6	0.713	0.688	0.694	0.725	0.719	0.002	0.008	0.002	0.006	0.004
0.65	0.726	0.688	0.686	0.732	0.718	0.004	0.017	0.007	0.002	0.006
0.7	0.73	0.714	0.716	0.728	0.726	0.01	0.004	0.002	0.002	0
0.75	0.739	0.721	0.723	0.726	0.726	0.004	0.015	0.006	0	0.001
0.8	0.735	0.723	0.725	0.723	0.724	0.004	0.001	0.001	0.004	0.002
0.85	0.728	0.722	0.726	0.717	0.72	0.013	0.004	0.001	0.006	0.007
0.9	0.728	0.721	0.722	0.719	0.715	0.013	0.002	0.001	0.006	0.004
0.95	0.727	0.715	0.715	0.717	0.715	0.013	0.004	0.005	0.004	0.004
1	0.727	0.715	0.715	0.715	0.715	0.012	0.004	0.004	0.004	0.004

(**b**)

	Gran_Size				
radius	***nil***	***A***	***B***	***C***	***D***
0	2	2	2	2	2
0.05	2	2	2	2.2	2.2
0.1	2.12	2	2	2.68	2.76
0.15	2.28	2.12	2.12	4.36	4.48
0.2	3.8	2.12	2.12	5.8	5.44
0.25	4.64	2.32	2.52	8.36	8.68
0.3	7.64	3.28	3.52	14	13.96
0.35	11.64	4.44	4.44	23.6	23.64
0.4	19.44	7	6.96	42.64	42.76
0.45	34.48	9.92	9.68	78.48	78.32
0.5	60.48	18.16	18.16	142.24	142.36
0.55	104.2	26.52	26.52	247.16	248.76
0.6	186.76	49.36	49.2	400.92	398.32
0.65	317.76	84.28	87.48	569.32	573.2
0.7	486.04	160.16	160	710.44	708.8
0.75	650.08	284.2	276.24	772.68	772.2
0.8	750.72	455.84	465.68	795.2	795.12
0.85	789.48	653.04	657.44	798.72	798.72
0.9	796.2	761	761.52	799.6	799.6
0.95	798.6	794.48	794.48	799.88	799.88
1	800	799.36	799.36	800	800

Table 7. Missing values handling using *conc_dep* homogeneous granulation technique; 5 × *Cross_V*5; *A*, *B*, *C*, *D* variants vs. nil case (classification based on original, undamaged training system); *Australian Credit*; *synthetic* 10% *damage*; *radius* = indiscernibility ratio; *Bias_Acc* = defined in Equation (1); *Gran_Size* = the number of training objects after granulation.

(**a**)

Accuracy					**Bias_Acc**				
nil	***A***	***B***	***C***	***D***	***nil***	***A***	***B***	***C***	***D***
0.842	0.847	0.85	0.848	0.845	0.003	0.001	0.001	0.001	0.006

(**b**)

Gran_Size				
nil	***A***	***B***	***C***	***D***
283.8	436.56	438	313.96	315.36

Table 8. Missing values handling using *conc_dep* homogeneous granulation technique; 5 × *Cross_V*5; *A*, *B*, *C*, *D* variants vs. nil case (classification based on original, undamaged training system); *Pima Indians Diabetes*; *synthetic* 10% *damage*; *radius* = indiscernibility ratio; *Bias_Acc* = defined in Equation (1); *Gran_Size* = the number of training objects after granulation.

(**a**)

Accuracy					**Bias_Acc**				
nil	***A***	***B***	***C***	***D***	***nil***	***A***	***B***	***C***	***D***
0.651	0.642	0.641	0.645	0.65	0.009	0.015	0.012	0.01	0.02

(**b**)

Gran_Size				
nil	***A***	***B***	***C***	***D***
487.52	578.52	579.56	489.72	493.2

Table 9. Missing values handling using *conc_dep* homogeneous granulation technique; 5 × *Cross_V*5; *A*, *B*, *C*, *D* variants vs. nil case (classification based on original, undamaged training system); *Heart Disease*; *synthetic* 10% *damage*; *radius* = indiscernibility ratio; *Bias_Acc* = defined in Equation (1); *Gran_Size* = the number of training objects after granulation.

(**a**)

Accuracy					**Bias_Acc**				
nil	***A***	***B***	***C***	***D***	***nil***	***A***	***B***	***C***	***D***
0.825	0.826	0.824	0.83	0.827	0.012	0.011	0.013	0.014	0.024

(**b**)

Gran_Size				
nil	***A***	***B***	***C***	***D***
120.48	159.08	157.96	127.16	126.84

Table 10. Missing values handling using *conc_dep* homogeneous granulation technique; 5 × *Cross_V*5; A, B, C, D variants vs. nil case (classification based on original, undamaged training system); *Hepatitis*; *synthetic* 10% *damage*; *radius* = indiscernibility ratio; *Bias_Acc* = defined in Equation (1); *Gran_Size* = the number of training objects after granulation.

(**a**)

Accuracy					**Bias_Acc**				
nil	***A***	***B***	***C***	***D***	***nil***	***A***	***B***	***C***	***D***
0.876	0.877	0.876	0.875	0.877	0.034	0.013	0.027	0.015	0.013

(**b**)

Gran_Size				
nil	***A***	***B***	***C***	***D***
45.76	57.12	57.68	50.92	51.4

Table 11. Missing values handling using *conc_dep* homogeneous granulation technique; 5 × *Cross_V*5; A, B, C, D variants vs. nil case (classification based on original, undamaged training system); *German credit*; *synthetic* 10% *damage*; *radius* = indiscernibility ratio; *Bias_Acc* = defined in Equation (1); *Gran_Size* = the number of training objects after granulation.

(**a**)

Accuracy					**Bias_Acc**				
nil	***A***	***B***	***C***	***D***	***nil***	***A***	***B***	***C***	***D***
0.726	0.72	0.718	0.733	0.728	0.007	0.004	0.013	0.002	0.007

(**b**)

Gran_Size				
nil	***A***	***B***	***C***	***D***
511.12	599.6	603.76	535.88	538.92

Comparing those results to the homogeneous granulation as a missing values absorption method, those gave the following findings. This technique is increasing the number of granules in the coverings—see Tables 7–11—and the indiscernability, in the context of decision classes, is lowering. This gives a higher probability of finding an object which breaks the homogeneity of the formed granule. Despite the fact that strategies A and B are returning smaller granules than in case C or D, the final granular reflection systems are bigger.

For given parameters, our methods work in a stable way, and the results are comparable to the *nil* case. A single run which is performed during the homogeneous granulation process is its biggest advantage, which might be the decisive factor when looking for the most robust method.

The results, showing our techniques using the strategy of completing unknown values with the most common values [31], can be found in Table 12. As we can see, they are equivalent to the results for the radius 1, in our strategies, where there is no approximation of training systems. Additionally, in Table 13, we have included degrees of homogeneity of the examined systems, i.e., the range of radii that appears during the homogeneous granulation process.

Table 12. Missing values handling using the most common value strategy; $5 \times Cross\_V5$; we consider repair options when $*$ =*each value*, $* = *$ and nil case (classification based on original, undamaged training system) *synthetic* 10% *damage*; *Bias_Acc* = defined in Equation (1); *Trn_Size* = Average number of training objects, d_1 = Australian Credit, d_2 = Pima Indians Diabetes, d_3 = Heart Disease, d_4 = Hepatitis, d_5 = German Credit.

	Accuracy			**Bias_Acc**			**Trn_Size**		
Data Set	***nil***	***∗ = Each Value***	***∗ = ∗***	***nil***	***∗ = Each Value***	***∗ = ∗***	***nil***	***∗ = Each Value***	***∗ = ∗***
d_1	0.862	0.849	0.849	0.012	0.008	0.008	552	550.56	550.56
d_2	0.651	0.636	0.636	0.006	0.014	0.014	614.4	613.48	613.48
d_3	0.829	0.83	0.83	0.008	0.007	0.007	216	215.64	215.64
d_4	0.892	0.884	0.884	0.005	0.019	0.019	124	123.36	123.36
d_5	0.727	0.715	0.715	0.012	0.004	0.004	800	799.36	799.36

Table 13. The degree of homogeneity—in the sense of homogeneous granulation—of the examined systems.

Name	***Radii_Range***
$Australian - credit$	$r_u \geq 0.5$
$Diabetes$	$r_u \geq 0.25$
$Heartdisease$	$r_u \geq 0.461$
$Hepatitis$	$r_u \geq 0.579$
$German - credit$	$r_u \geq 0.6$

4. Conclusions

Comparing concept dependent and homogeneous granulation as a missing values absorption technique, we can point to the following conclusions.

The $* = each\ value$ variant used with concept dependent granulation generates more approximate datasets (diversity reduction) while the $* = *$ case may increase the diversity. The granules are smaller for C and D strategies compared to the strategies A and B. Granulation of systems containing missing values reduces its size to a much higher degree than the granulation of undamaged datasets.

We can observe specific results when using homogeneous granulation as a missing values absorption technique. When comparing the results to the nil case—granulation of the undamaged dataset—granules in A and B strategies are smaller than those from C and D. It is happening because the $* = each\ value$ case is breaking the homogeneity of the decision classes to a higher degree than the $* = *$ case. The approximation level is decreasing for damaged datasets.

Granulation techniques are absorbing missing values in an effective way as confirmed by the classification results of the $Cross\_V$ model. The most missing values are repaired during the granulation process no matter which technique is being used.

In our research, we are going to choose the most effective technique among known classifiers for specific types of data. We also plan to implement and check effectiveness of homogeneous granulation in the context of classification based on deep neural networks.

Author Contributions: Conceptualization, P.A. and K.R.; Methodology, P.A. and K.R.; Software, P.A. and K.R.; Validation, P.A. and K.R.; Formal Analysis, P.A. and K.R.; Investigation, P.A. and K.R.; Resources, P.A. and K.R.; Writing—Original Draft Preparation, P.A. and K.R.; Writing—Review and Editing, P.A. and K.R.; Visualization, P.A. and K.R.; Project Administration, P.A. and K.R. Funding Acquisition, P.A. and K.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been fully supported by the grant from the Ministry of Science and Higher Education of the Republic of Poland under the project number 23.610.007-000.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Zadeh, L.A. Fuzzy sets and information granularity. *Adv. Fuzzy Set Theory Appl.* **1979**, *11*, 3–18.
2. Zadeh, L.A. Graduation and granulation are keys to computation with information described in natural language. In Proceedings of the 2006 IEEE International Conference on Granular Computing, Atlanta, GA, USA, 10–12 May 2006; p. 30.
3. Skowron, A.; Polkowski, L. Synthesis of decision systems from data tables. In *Rough Sets and Data Mining*; Lin, T.Y.; Cercone, N., Eds.; Kluwer: Dordrecht, The Netherlands, 1997; pp. 289–299.
4. Lin, T.Y. Granular computing: Examples, intuitions and modeling. In Proceedings of the 2005 IEEE International Conference on Granular Computing, Beijing, China, 25–27 July 2005; Volume 1, pp. 40–44.
5. Yao, Y.Y. Granular computing: Basic issues and possible solutions. In *Proceedings 5th Joint Conference Information Sciences I*; Wang, P.P., Ed.; Association for Intellectual Machinery: Atlantic, NJ, USA, 2000; pp. 186–189.
6. Yao, Y. Information Granulation and Approximation in a Decision-Theoretical Model of Rough Sets. In *Rough-Neural Computing. Cognitive Technologies*; Pal, S.K., Polkowski, L., Skowron, A., Eds.; Springer: Berlin, Germany, 2004.
7. Yiyu, Y. Perspectives of granular computing. In Proceedings of the 2005 IEEE International Conference on Granular Computing, Beijing, China, 25–27 July 2005; Volume 1, pp. 85–90.
8. Skowron, A.; Stepaniuk, J. Information granules: Towards foundations of granular computing. *Int. J. Intell. Syst.* **2001**, *16*, 57–85. [CrossRef]
9. Skowron, A.; Stepaniuk, J. Information Granules and Rough-Neural Computing. In *Rough-Neural Computing. Cognitive Technologies*; Pal, S.K., Polkowski, L., Skowron, A., Eds.; Springer: Berlin, Germany, 2004; pp. 43–84.
10. Polkowski, L.; Semeniuk–Polkowska, M. On rough set logics based on similarity relations. *Fund. Inf.* **2005**, *64*, 379–390.
11. Liu, Q.; Sun, H. Theoretical study of granular computing. In *Proceedings RSKT06, Chongqing, China, 2006—Lecture Notes in Artificial Intelligence 4062*; Springer: Berlin, Germany 2006; pp. 92–102.
12. Cabrerizo, F.J.; Al-Hmouz, R.; Morfeq, A.; Martínez, M.A.; Pedrycz, W.; Herrera-Viedma, E. Estimating incomplete information in group decision-making: A framework of granular computing. *Appl. Soft Comput.* **2020**, *86*, 105930. [CrossRef]
13. Capizzi, G.; Lo Sciuto, G.; Napoli, C.; Połap, D.; Woźniak, M. Small Lung Nodules Detection based on Fuzzy-Logic and Probabilistic Neural Network with Bio-inspired Reinforcement Learning. *IEEE Trans. Fuzzy Syst.* **2019**. Available online: https://ieeexplore.ieee.org/abstract/document/8895990 (accessed on 13 February 2020).
14. Hryniewicz, O.; Kaczmarek, K. Bayesian analysis of time series using granular computing approach. *Appl. Soft Comput.* **2016**, *47*, 644–652. [CrossRef]
15. Martino, A.; Giuliani, A.; Rizzi, A. (Hyper) Graph Embedding and Classification via Simplicial Complexes. *Algorithms* **2019**, *12*, 223. [CrossRef]
16. Martino, A.; Giuliani, A.; Todde, V.; Bizzarri, M.; Rizzi, A. Metabolic networks classification and knowledge discovery by information granulation. *Comput. Biol. Chem.* **2020**, *84*, 107187. [CrossRef] [PubMed]
17. Pownuk, A.; Kreinovich, V. Granular approach to data processing under probabilistic uncertainty. In *Granular Computing*; Springer: Berlin, Germany, 2019; pp. 1–17.
18. Zhong, C.; Pedrycz, W.; Wang, D.; Li, L.; Li, Z. Granular data imputation: A framework of granular computing. *Appl. Soft Comput.* **2016**, *46*, 307–316. [CrossRef]
19. Grzymala-Busse J.W.; Grzymala-Busse W.J. Handling Missing Attribute Values. In *Data Mining and Knowledge Discovery Handbook*; Maimon O., Rokach L., Eds.; Springer: Boston, MA, USA, 2005.
20. Polkowski, L. A model of granular computing with applications. In Proceedings of the IEEE 2006 Conference on Granular Computing GrC06, Atlanta, GA, USA, 10–12 May 2006; pp. 9–16.
21. Polkowski, L. Formal granular calculi based on rough inclusions. In Proceedings of the IEEE 2005 Conference on Granular Computing GrC05, Beijing, China, 25–27 July 2005; pp. 57–62.
22. Polkowski, L.; Artiemjew, P. Granular Computing in Decision Approximation—An Application of Rough Mereology. In *Intelligent Systems Reference Library 77*; Springer: Berlin, Germany, 2015; pp. 1–422, ISBN 978-3-319-12879-5.

23. Ropiak, K.; Artiemjew, P. On Granular Rough Computing: Epsilon homogenous granulation. In *Proceedings of International Joint Conference on Rough Sets, IJCRS'18, Quy Nhon, Vietnam, Lecture Notes in Computer Science (LNCS)*; Springer: Berlin, Germany, 2018.
24. Ropiak, K.; Artiemjew, P. A Study in Granular Computing: Homogenous granulation. In *Proceedings of the Information and Software Technologies—ICIST 2018—Communications in Computer and Information Science*; Dregvaite, G., Damasevicius, R., Eds.; Springer: Berlin, Germany, 2018.
25. Artiemjew, P.; Ropiak, K. Missing Values Absorption Based on Homogenous Granulation. In *Information and Software Technologies—ICIST 2019—Communications in Computer and Information Science*; Damaševičius, R., Vasiljeviene, G., Eds.; Springer: Berlin, Germany, 2019; Volume 1078.
26. Ropiak, K.; Artiemjew, P. Homogenous Granulation and Its Epsilon Variant. *Computers* **2019**, *8*, 36. [CrossRef]
27. Artiemjew, P.; Ropiak, K. A Novel Ensemble Model—The Random Granular Reflections. In Proceedings of the 27th International Workshop on Concurrency, Specification and Programming, CEUR, Berlin, Germany, 24–26 September 2018.
28. Polkowski, L.; Artiemjew, P. On Granular rough computing with missing values. In *Proceedings of the International Conference on Rough Sets and Intelligent Systems Paradigms RSEiSP'07, Lecture Notes in Computer Science*; Springer: Berlin, Germany, 2007; Volume 4585, pp. 271–279.
29. Polkowski, L.; Artiemjew, P. Granular computing: Granular classifiers and missing values. In Proceedings of the 6th IEEE International Conference on Cognitive Informatics ICCI'07, Lake Tahoo, CA, USA, 6–8 August 2007; pp. 186–194.
30. UCI Data Repository: Available online: https://archive.ics.uci.edu/ml/index.php (accessed on 13 February 2020).
31. Jerez, J.M.; Molina, I.; García-Laencina, P.J.; Alba, E.; Ribelles, N.; Martín, M.; Franco, L. Missing data imputation using statistical and machine learning methods in a real breast cancer problem. *Artif. Intell. Med.* **2010**, *50*, 105–115. [CrossRef] [PubMed]